GENERIC IDENTIFIABILITY OF LINEAR STRUCTURAL EQUATION MODELS BY ANCESTOR DECOMPOSITION

MATHIAS DRTON AND LUCA WEIHS

ABSTRACT. Linear structural equation models, which relate random variables via linear interdependencies and Gaussian noise, are a popular tool for modeling multivariate joint distributions. These models correspond to mixed graphs that include both directed and bidirected edges representing the linear relationships and correlations between noise terms, respectively. A question of interest for these models is that of parameter identifiability, whether or not it is possible to recover edge coefficients from the joint covariance matrix of the random variables. For the problem of determining generic parameter identifiability, we present an algorithm that extends an algorithm from prior work by Foygel, Draisma, and Drton (2012). The main idea underlying our new algorithm is the use of ancestral subsets of vertices in the graph in application of a decomposition idea of Tian (2005).

1. INTRODUCTION

It is often useful to model the joint distribution of a random vector $X = (X_1, ..., X_n)^T$ in terms of a collection of noisy linear interdependencies. In particular, we may postulate that each X_w is a linear function of $X_1, ..., X_{w-1}, X_{w+1}, ..., X_n$ and a stochastic noise term ϵ_w. Models of this type are called *linear structural equation models* and can be compactly expressed in matrix form as

$$(1.1) \qquad X = \lambda_0 + \Lambda^T X + \epsilon$$

where $\Lambda = (\lambda_{vw})$ is a $n \times n$ matrix, $\lambda_0 = (\lambda_{01}, ..., \lambda_{0n})^T \in \mathbb{R}^n$, and $\epsilon = (\epsilon_1, ..., \epsilon_n)^T$ is a random vector of error terms. We will adopt the classical assumption that ϵ has a non-degenerate multivariate normal distribution with mean 0 and covariance matrix $\Omega = (\omega_{vw})$. With this assumption it follows immediately that X has a multivariate normal distribution with mean $(I - \Lambda)^{-T} \lambda_0$ and covariance matrix

$$(1.2) \qquad \Sigma = (I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1}$$

where I is the $n \times n$ identity matrix. We refer the reader to the book by Bollen (1989) for background on these types of models.

We obtain a collection of interesting models by imposing different patterns of zeros among the coefficients in Λ and Ω. These models can then be naturally associated with mixed graphs containing both directed and bidirected edges. In particular, the graph will contain the directed edge $v \to w$ when λ_{vw} is not required to be zero and, similarly, will include the bidirected edge $v \leftrightarrow w$ when ω_{vw} is potentially non-zero. Representations of this type are often called *path diagrams* and were first advocated for in Wright (1921, 1934).

A natural question arising in the study of linear structural equations is that of identifiability; whether or not it is possible to uniquely recover the two parameter matrices Λ and Ω from the covariance matrix Σ they define via (1.2). The most stringent version, known as global identifiability, amounts to unique recovery of every pair (Λ, Ω) from the covariance matrix Σ. This global property can be characterized efficiently (Drton et al., 2011). Often, however, a less stringent notion that we term generic identifiability is of interest. This property requires only that a generic (or randomly chosen) pair (Λ, Ω) can be recovered from its covariance matrix. The computational complexity of deciding whether a given mixed graph G defines a generically identifiable linear structural equation model is unknown. There are, however, a number of graphical criteria that are sufficient for generic identifiability and can be checked in polynomial time in the number of considered variables (or vertices of the graph). To our knowledge, the most widely applicable such criterion is the Half-Trek Criterion (HTC) of Foygel, Draisma, and Drton (2012a), which built on earlier work of Brito and Pearl (2006). The HTC also comes with a necessary condition for generic identifiability but in this paper our focus is on the sufficient condition. We remark that an extension of the HTC for identification of subsets of edge coefficients is given in Chen et al. (2014).

We begin with a brief review of background such as the formal connection between structural equation models and mixed graphs, and give a review of prior work in Section 2. In the main Section 3, we will demonstrate a simple method by which to infer generic identifiability of certain entries of (Λ, Ω) by examining subgraphs of a given mixed graph G that are induced by ancestral subsets of vertices. This will extend the applicability of the HTC in the case of acyclic mixed graphs after applying the decomposition techniques of Tian (2005). We leverage this extension in an efficient algorithmic form. In Section 4, we report on computational experiments demonstrating the applicability of our findings. A brief conclusion is given in Section 5.

2. Preliminaries

We assume that the reader is familiar with graphical representations of structural equation models and thus only provide a quick review of these topics. For a more in-depth treatment see, for instance, Pearl (2009) or Foygel et al. (2012a).

2.1. **Mixed Graphs.** For any $n \geq 1$, let $[n] := \{1, ..., n\}$. We define a *mixed graph* to be a triple $G = (V, D, B)$ where $V = [n]$ is a finite set of vertices and $D, B \subset V \times V$. The sets D and B correspond to the directed and the bidirected edges, respectively. When $(v, w) \in D$, we will write $v \to w \in G$ and if $(v, w) \in B$ then we will write $v \leftrightarrow w \in G$. Since edges in B are bidirected the set B is symmetric, that is, we have $(v, w) \in B \iff (w, v) \in B$. We require that both the directed part (V, D) and bidirected part (V, B) contain no self loops so that $(v, v) \notin D \cup B$ for all $v \in V$. If the directed graph (V, D) does not contain any cycles, so that there are no vertices $v, w_1, ..., w_m \in V$ such that $v \to w_1, w_1 \to w_2, ..., w_m \to v \in G$, then we say that G is *acyclic*; note, in particular, that G being acyclic does not imply (V, B) does not contain any (undirected) cycles.

A *path* from v to w is any sequence of edges from D or B beginning at v and ending with w, here edges need not obey direction and loops are allowed. A *directed path* from v to w is then any path from v to w all of whose edges are directed and pointed in the same direction, away from v and towards w. Finally, a *trek* π from

a *source* v to a *target* w is any path that has no colliding arrowheads, that is, π must be of the form

$$v_l^L \leftarrow v_{l-1}^L \leftarrow ... \leftarrow v_0^L \longleftrightarrow v_0^R \rightarrow v_1^R \rightarrow ... \rightarrow v_{r-1}^R \rightarrow v_r^R$$

or

$$v_l^L \leftarrow v_{l-1}^L \leftarrow ... \leftarrow v_1^L \leftarrow v^T \rightarrow v_1^R \rightarrow ... \rightarrow v_r^R \rightarrow v_r^R$$

where $v_l^L = v$, $v_r^R = w$, and we call v^T the *top* node. If π is as in the first case then we let $\text{Left}(\pi) = \{v_0^L, ..., v_l^L\}$ and $\text{Right}(\pi) = \{v_0^R, ..., v_r^R\}$, if π is as in the second case then we let $\text{Left}(\pi) = \{v^T, v_1^L, ..., v_l^L\}$ and $\text{Right}(\pi) = \{v^T, v_1^R, ..., v_r^R\}$. Note that, in the second case, v^T is included in both $\text{Left}(\pi)$ and $\text{Right}(\pi)$. A trek π is called a *half-trek* if $|\text{Left}(\pi)| = 1$ so that π is of the form

$$v_0^L \longleftrightarrow v_0^R \rightarrow v_1^R \rightarrow ... \rightarrow v_{r-1}^R \rightarrow v_r^R$$

or

$$v^T \rightarrow v_1^R \rightarrow ... \rightarrow v_r^R \rightarrow v_r^R$$

It will be useful to reference the local neighborhood structure of the graph. For this purpose, for all $v \in V$, we define the two sets

$$(2.1) \qquad\qquad pa(v) = \{w \in V : w \rightarrow v \in G\},$$

$$(2.2) \qquad\qquad sib(v) = \{w \in V : w \leftrightarrow v \in G\}.$$

The former comprises the *parents* of v, and the latter contains the *siblings* of v.

We associate a mixed graph G to a linear structural equation model as follows. Let $\mathbb{R}^D$ be the set of real $n \times n$ matrices $\Lambda = (\lambda_{vw})$ with support D, i.e., $\lambda_{vw} \neq 0 \implies (v, w) \in D$. Let PD_n be the cone of $n \times n$ positive definite matrices $\Omega = (\omega_{vw})$. Define $PD(B) \subset PD_n$ to be the subset of positive definite matrices with support B, i.e. for $v \neq w$, $\omega_{vw} \neq 0 \implies v \leftrightarrow w \in G$.

In this paper, we focus on acyclic graphs G. If G is acyclic then the matrix $I - \Lambda$ is invertible for all $\Lambda \in \mathbb{R}^D$. In other words, the equation system from (1.1) can always be solved uniquely for X. We are led to the following definition.

Definition 2.1. The *linear structural equation model given by an acyclic mixed graph* $G = (V, D, B)$ with $V = [n]$ is the collection of all n-dimensional normal distributions with covariance matrix

$$\Sigma = (I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1}$$

for a choice of $\Lambda \in \mathbb{R}^D$ and $\Omega \in PD(B)$.

2.2. **Prior Work and the HTC.** For a fixed acyclic mixed graph G, let $\Theta := \mathbb{R}^D \times PD(B)$ be the parameter space and $\phi_G : \Theta \rightarrow PD_n$ be the map

$$(2.3) \qquad\qquad \phi_G : (\Lambda, \Omega) \mapsto (I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1}.$$

Then the question of identifiability is equivalent to asking whether the *fiber*

$$\mathcal{F}(\Lambda, \Omega) := \phi_G^{-1}(\{\phi_G(\Lambda, \Omega)\})$$

equals the singleton $\{(\Lambda, \Omega)\}$. We note that the above notions are well-defined also when G is not acyclic but, in that case, $\mathbb{R}^D$ should be restricted to contain only matrices Λ with $I - \Lambda$ invertible.

When $\mathcal{F}(\Lambda, \Omega) = \{(\Lambda, \Omega)\}$ for all $(\Lambda, \Omega) \in \Theta$, so that ϕ_G is injective on Θ, then G is said to be *globally identifiable*. Global identifiability is, however, often too

strong a condition. So-called instrumental variable problems, for instance, give rise to graphs G that would not be globally identifiable but for which the set of (Λ, Ω) on which identifiability fails has measure zero; see the example in the introduction of Foygel et al. (2012a). Instead, we will be concerned with the question of generic identifiability.

Definition 2.2. A mixed graph G is said to be *generically identifiable* if there exists a proper algebraic subset $A \subset \Theta$ so that G is identifiable on $\Theta \setminus A$.

Here, as usual, an *algebraic set* is defined as the zero-set of a collection of polynomials. We again refer the reader to the introduction of Foygel et al. (2012a) for an in-depth exposition on why generic identifiability is an often appropriate weakening of global identifiability.

Now there will be cases when we are interested in understanding the generic identifiability of certain coefficients of a mixed graph G rather than all coefficients simultaneously. In these cases we say that the coefficient λ_{vu} (or ω_{vu}), for $u, v \in V$, is generically identifiable in G if the projection of the fiber $\mathcal{F}(\Lambda, \Omega)$ onto λ_{vu} (or ω_{vu}) is a singleton for all $\Theta \setminus A$ where $A \subset \Theta$ is a proper algebraic set.

Let Λ and Ω be matrices of indeterminates as in Equation (1.2) with zero pattern corresponding to G. Then, by the Trek Rule of Wright (1921), see also Spirtes, Glymour, and Scheines (2000), the covariance Σ_{vw} can be represented as a sum of monomials corresponding to treks between v and w in G. To state the Trek Rule formally, let $\mathcal{T}(v, w)$ be the set of all treks from v to w in G. Then for any $\pi \in \mathcal{T}(v, w)$, if π contains no bidirected edge and has top node z, we define the *trek monomial* as

$$\pi(\Lambda, \Omega) = \omega_{zz} \prod_{x \to y \in \pi} \lambda_{xy},$$

and if π contains a bidirected edge connecting $u, z \in V$ then we define the trek monomial as

$$\pi(\Lambda, \Omega) = \omega_{uz} \prod_{x \to y \in \pi} \lambda_{xy}$$

We may then state the rule as follows.

Proposition 2.3 (Trek Rule). *For all $v, w \in V$, the covariance matrix $\Sigma = (I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1}$ corresponding to a mixed graph G satisfies*

$$\Sigma_{vw} = \sum_{\pi \in \mathcal{T}(v,w)} \pi(\Lambda, \Omega).$$

Before giving the statement of the HTC, we must first define what is meant by a half-trek system. Let $\Pi = \{\pi_1, ..., \pi_m\}$ be a collection of m treks with each π_i having source x_i and target y_i, then Π is called a *system of treks* from $X = \{x_1, ..., x_m\}$ to $Y = \{y_1, ..., y_m\}$ if $|X| = |Y| = m$ so that all sources as well as all targets are pairwise distinct. If each π_i is a half-trek, then Π is a *system of half-treks*. Moreover, a collection $\Pi = \{\pi_1, ..., \pi_m\}$ of treks is said to have *no sided intersection* if

$$\text{Left}(\pi_i) \cap \text{Left}(\pi_j) = \varnothing = \text{Right}(\pi_i) \cap \text{Right}(\pi_j), \ \forall i \neq j$$

Let $htr(v)$ be the collection of vertices $w \in V \setminus (\{v\} \cup sib(v))$ for which there is a half-trek from v to w, these w are called *half-trek reachable* from v. We have the following definition and result of Foygel et al. (2012a).

Definition 2.4. A set of nodes $Y \subset V$ satisfies the *half-trek criterion* with respect to a node $v \in V$ if

(i) $|Y| = |pa(v)|$,

(ii) $Y \cap (\{v\} \cup sib(v)) = \varnothing$, and

(iii) there is a system of half-treks with no sided intersection from Y to $pa(v)$.

Theorem 2.5 (HTC-identifiability). *Let $(Y_v : v \in V)$ be a family of subsets of the vertex set V of a mixed graph G. If, for each node v, the set Y_v satisfies the half-trek criterion with respect to v, and there is a total ordering $\prec$ on the vertex set V such that $w \prec v$ whenever $w \in Y_v \cap htr(v)$, then G is rationally identifiable.*

The assertion that G is *rationally identifiable* means that the inverse map ϕ_G^{-1} can be represented as a rational function on $\Theta \setminus A$ where A is some proper algebraic subset of Θ. Clearly, rational identifiability is a stronger condition than generic identifiability. If a graph G satisfies Theorem 2.5 we will say that G is *HTC-identifiable* (HTCI). In a similar vein, Theorem 2 of Foygel et al. (2012a) gives sufficient conditions for a graph G to be generically unidentifiable (with generically infinite fibers of ϕ_G), and we will call such graphs *HTC-unidentifiable* (HTCU). Graphs that are neither HTCI or HTCU are called *HTC-inconclusive*, these are the graphs on which progress is left to be made.

As is noted in Section 8 in Foygel et al. (2012a) we may extend the power of the HTC by using the graph decomposition techniques of Tian (2005). Let $C_1, ..., C_k \subset V$ be the unique partitioning of V where $v, w \in C_i$ if and only if there exists a (possibly empty) path from v to w composed of only bidirected edges. In other words, $C_1, ..., C_k$ are the connected components of (V, B), the bidirected part of G. For $i = 1, \ldots, k$, let

$$V_i = C_i \cup pa(C_i), \qquad\qquad D_i = \{v \to w \in G : v \in V_i,\ w \in C_i\},$$
$$B_i = \{v \leftrightarrow w \in G : v, w \in C_i\}, \qquad G_i = (V_i, D_i, B_i).$$

Then the mixed graphs $G_1, ..., G_k$ are called the *mixed components* of G. From the work of Tian (2005), Foygel et al. (2012a) present the following theorem.

Theorem 2.6 (Tian Decomposition). *For an acyclic mixed graph G with mixed components $G_1, ..., G_k$, the following holds:*

(i) *G is rationally (or generically) identifiable if and only if all components $G_1, ..., G_k$ are rationally (or generically) identifiable;*

(ii) *G is generically infinite-to-one if and only if there exists a component G_j that is generically infinite-to-one;*

(iii) *if each G_j is generically h_j-to-one with $h_j < \infty$, then G is generically h-to-one with $h = \prod_{j=1}^{k} h_j$.*

We remark that this decomposition also plays a role in non-linear models; see, for instance, the paper of Shpitser et al. (2014) and the references given therein.

3. Ancestral Decomposition

For a later strengthening of the HTC, we will show that the generic identification of certain subgraphs of an acyclic mixed graph $G = (V, D, B)$ implies the generic identification of their associated edge coefficients in the larger graph G. This result is straightforward and is well known in other forms. Surprisingly, however, this

simple idea can extend the applicability of the HTC when combined with the decomposition from Theorem 2.6. We will first define what we mean by an ancestral subset and an induced graph,

Definition 3.1. Let $V' \subset V$ be a subset of vertices. The *ancestors of V'* form the set

$$An(V') = \{v \in V : \text{there exists a directed path from } v \text{ to some } w \in V'\},$$

where we consider the empty path to be directed so that $V' \subset An(V')$. If $V' = An(V')$, then we call V' *ancestral*.

Definition 3.2. Let $V' \subset V$ be again a subset of vertices. The *subgraph of G induced by V'* is the mixed graph $G_{V'} = (V', D', B')$ with

$$D' = \{v \to w \in G : v, w \in V'\},$$
$$B' = \{v \leftrightarrow w \in G : v, w \in V'\}.$$

We now have the following simple fact.

Theorem 3.3. *Let $G = (V, D, B)$ be a mixed graph, and let V' be an ancestral subset of V. If the induced subgraph $G_{V'}$ is generically (or rationally) identifiable then so are all the corresponding edge coefficients in G.*

Proof. Let the covariance matrix $\Sigma = (I - \Lambda)^{-T} \Omega (I - \Lambda)^{-1}$ correspond to G, that is, $\Lambda \in \mathbb{R}^D$ and $\Omega \in PD(B)$. Let Λ' and Ω' denote the $V' \times V'$ submatrices of Λ and Ω, respectively, and let $\Sigma' = (I_{|V'|} - \Lambda')^{-T} \Omega (I_{|V'|} - \Lambda')^{-1}$ where $I_{|V'|}$ is the $|V'| \times |V'|$ identity matrix. For ease of notation, write $G' = G_{V'}$.

Recall that for any $v, w \in V$, the set $\mathcal{T}(v, w)$ comprises all treks between v and w in G. Similarly, write $\mathcal{T}_{G'}(v, w)$ for the set of treks between v and w in G'. Since V' is ancestral, it holds that $\mathcal{T}(v, w) = \mathcal{T}_{G'}(v, w)$ for all $v, w \in V'$. Thus, by Proposition 2.3, we have that for any $v, w \in V'$

$$\Sigma_{vw} = \sum_{\pi \in \mathcal{T}(v,w)} \pi(\Lambda, \Omega) = \sum_{\pi \in \mathcal{T}_{G'}(v,w)} \pi(\Lambda', \Omega') = \Sigma'_{vw}.$$

Now suppose that G' is generically (or rationally) identifiable. Then Λ', Ω' can be generically (or rationally) recovered from Σ'. As we have just shown that $\Sigma_{vw} = \Sigma'_{vw}$ for all $v, w \in V'$, we have that the entries of Λ, Ω corresponding to Λ', Ω' can be recovered from Σ generically (or rationally). $\square$

We may generalize the above theorem so that we do not have to consider the identifiability of all of G' and instead only look at certain edges in G'.

Corollary 3.4. *Let $G = (V, D, B)$ be a mixed graph, and let V' be an ancestral subset of V. If an edge coefficient of $G_{V'}$ is generically (or rationally) identifiable then so is the corresponding coefficient in G.*

Proof. This follows exactly as in the proof of Theorem 3.3 by only considering a single generically (or rationally) identifiable coefficient of G' at a time. $\square$

We give an example as to how Theorem 3.3 strengthens the HTC.

Example 3.5. It is straightforward to check that the graph G from Figure 1a is HTC-inconclusive using Algorithm 1 from Foygel et al. (2012a). We direct the reader who does not want to perform this computation by hand to the R package
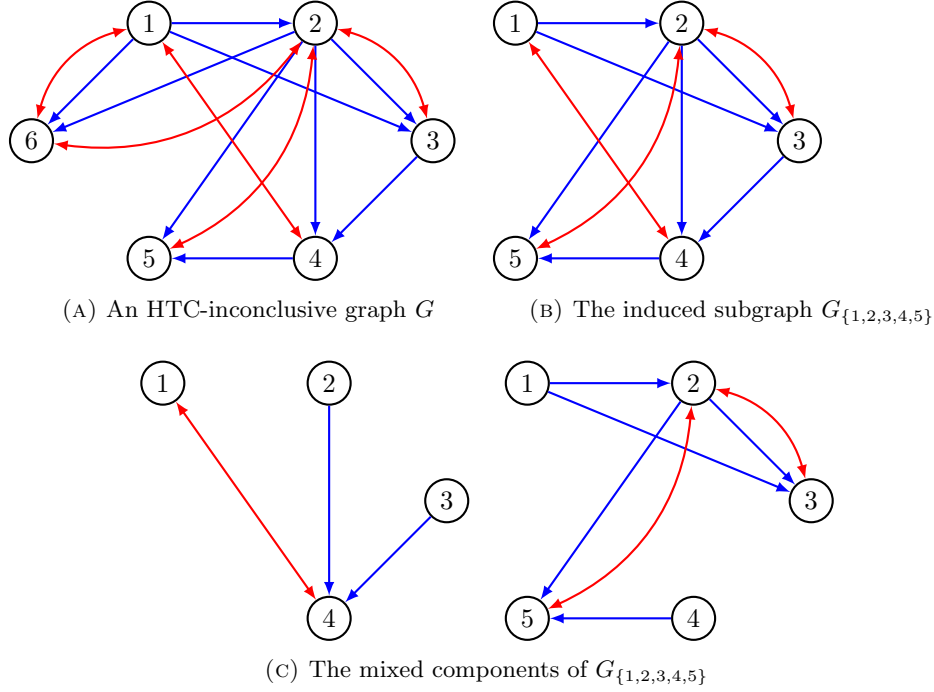
(A) An HTC-inconclusive graph G

(B) The induced subgraph $G_{\{1,2,3,4,5\}}$

(C) The mixed components of $G_{\{1,2,3,4,5\}}$

FIGURE 1. A mixed graph G, a subgraph induced by an ancestral subset, and the mixed components the induced graph.

SEMID (Foygel and Drton, 2013; R Core Team, 2014). Moreover, G cannot be decomposed as its bidirected part is connected.

Now the set $V' = \{1, 2, 3, 4, 5\}$ is ancestral in G, so we may apply Theorem 3.3 to the induced subgraph $G' = G_{1,2,3,4,5}$. While G' remains HTC-inconclusive, the Tian decomposition of Theorem 2.6 is applicable. After decomposing G' into its mixed components, see Figure 1c, we find that each component is HTC-identifiable and thus G' itself is generically identifiable. To show generic identifiability of G, we are left to show that all the coefficients on the directed edges between $pa(6) = \{1, 2\}$ and 6 are generically identifiable. Since $Y = \{3, 4\}$ satisfies the HTC with respect to 6 it follows, by Lemma 3.6 below, that $\Lambda_{pa(v),v}$ is generically identifiable. Hence, the entire matrix Λ is generically identifiable, and since $(I - \Lambda)^T \Sigma (I - \Lambda) = \Omega$, this implies generic identifiability of Ω. We conclude that G is generically identifiable despite being HTC-inconclusive.

Lemma 3.6. *Let $G = (V, D, B)$ be a mixed graph, and let $v \in V$. If $Y \subset V$ satisfies the HTC with respect to v and for each $y \in Y$ we have that $\Lambda_{pa(y),y}$ is generically (or rationally) identifiable, then $\Lambda_{pa(v),v}$ is generically (or rationally) identifiable.*

Proof. Suppose vertex v has m parents, and let $pa(v) = \{p_1, ..., p_m\}$. Since Y satisfies the HTC for v, we must have $|Y| = |pa(v)| = m$. Thus we may enumerate the set as $Y = \{y_1, ..., y_m\}$. Define a matrix $A \in \mathbb{R}^{m \times m}$ with entries

$$A_{ij} := \begin{cases} [(I - \Lambda)^T \Sigma]_{y_i p_j} & \text{if } y_i \in htr(v), \\ \Sigma_{y_i p_j} & \text{if } y_i \notin htr(v), \end{cases}$$

and define a vector $b \in \mathbb{R}^m$ with entries

$$b_i := \begin{cases} [(I - \Lambda)^T \Sigma]_{y_i v} & \text{if } y_i \in htr(v), \\ \Sigma_{y_i v} & \text{if } y_i \notin htr(v). \end{cases}$$

Both A and b are generically identifiable because we have assumed that $\Lambda_{pa(y),y}$ is generically identifiable for every $y \in Y$. Now, from the proof of Theorem 1 in Foygel et al. (2012a), we have $A \cdot \Lambda_{pa(v),v} = b$, and from Lemma 2 of Foygel et al. (2012a) we deduce that A is generically invertible. It follows that $\Lambda_{pa(v),v} = A^{-1}b$ generically so that $\Lambda_{pa(v),v}$ is generically identifiable.   □

Algorithm 1 from Foygel et al. (2012a) (hereafter called the HTC-algorithm) determines whether or not a mixed graph $G = (V, D, B)$ satisfies the conditions of Theorem 2.5 and and thus checks if the graph is HTCI. The HTC-algorithm operates by iteratively looping through the nodes $v \in V$ and attempting to find a half-trek system Π to $pa(v)$ with Π having sources which are in a set of "allowed" nodes $A \subset V$. Here, a node w is allowed if $w \notin htr(v) \cup \{v\} \cup sib(v)$ or if w was previously shown to be generically identifiable in the sense that all coefficients on directed edges $u \to w$ were shown to be generically identifiable. If such a half-trek system Π is found for node v then Foygel et al. (2012a) show that this implies that v is generically identifiable, and thus v may be added to the set of allowed nodes for the remaining iterations. The algorithm terminates when all nodes have been shown to be generically identifiable or once it has iterated through all vertices and has been unable to show the generic identifiability of any new nodes. To find a half-trek system between a suitable subset of the set of allowed nodes A and $pa(v)$, the HTC-algorithm solves a Max Flow problem on an auxiliary network $G_{\text{flow}}(A, v)$, and this step takes $\mathcal{O}(|V|^3)$ time when G is acyclic. If in $G_{\text{flow}}(A, v)$ one can find a flow of size $|pa(v)|$ then the half-trek system exists. See Section 6 of Foygel et al. (2012a) for more details about how $G_{\text{flow}}(A, v)$ is defined. Finally, for an acyclic mixed graph G, the HTC-algorithm has a worst case running time of $\mathcal{O}(|V|^5)$.

Algorithm 1 presents a simple modification of the HTC-algorithm to leverage Corollary 3.4 extending the ability of the HTC to determine the generic identifiability of acyclic mixed graphs. We emphasize that this algorithm considers only certain ancestral subsets and, as such, we do not necessarily expect the algorithm to reach a conclusion in all cases in which Corollary 3.4 may be applicable.

Proposition 3.7. *Algorithm 1 returns "yes" only if the input acyclic mixed graph G is generically identifiable and will return "yes" whenever the HTC-algorithm does. Moreover, Algorithm 1 returns "yes" for the HTC-inconclusive graph in Figure 1a and has time complexity at most $\mathcal{O}(|V|^5)$.*

Proof. The fact that Algorithm 1 only returns "yes" if G is generically identifiable follows from Theorem 7 in Foygel et al. (2012a,b) and our Corollary 3.4. That Algorithm 1 returns "yes" whenever the HTC-algorithm does can be argued as follows:

If, for a set of allowed nodes A and $v \in V$, there exists $Y \subset A$ satisfying the HTC for v then we must have that $Y \subset S := An(v) \cup sib(An(v))$ and Y satisfies the HTC for v in $G_{An(\{v\} \cup (A \cap S))}$. Lemma 4 of Foygel et al. (2012b) then yields that there exists a set Y' of allowed nodes which satisfies the HTC for v in the mixed component of $G_{An(\{v\} \cup (A \cap S))}$ containing v. Hence, if v is added to the set of solved nodes in the HTC-algorithm it will also be added to the set of solved nodes

---

Algorithm 1 A sufficient test for generic identifiability.

1: **Input:** $G = (V, D, B)$, an acyclic mixed graph on n nodes $v_1, \ldots, v_n$
2: **Initialize:** SolvedNodes $\leftarrow \{v : pa(v) = \varnothing\}$.
3: **repeat**
4:     **for** $v = v_1, v_2, \ldots, v_n$ **do**
5:         **if** $v \notin$ SolvedNodes **then**
6:             $\triangleright$ Check if we can generically identify $\Lambda_{pa(v),v}$ using
7:             $\triangleright$ the induced graph $G_{An(\{v\} \cup (A \cap S))}$.
8:             $S \leftarrow An(v) \cup sib(An(v))$
9:             $A \leftarrow S \cap (\text{SolvedNodes} \cup (V \setminus htr(v))) \setminus (\{v\} \cup sib(v))$
10:           $G' \leftarrow$ the mixed component of $G_{An(\{v\} \cup (A \cap S))}$ containing v
11:           $A \leftarrow (A \cap (\text{vertices in } G')) \cup (\text{source nodes in } G')$
12:           **if** $\text{MaxFlow}(G_{\text{flow}}(v, A)) = |pa(v)|$ **then**
13:              SolvedNode $\leftarrow$ SolvedNodes $\cup \{v\}$.
14:              Skip to next iteration of the loop
15:           **end if**
16:             $\triangleright$ Check if we can generically identify $\Lambda_{pa(v),v}$ using
17:             $\triangleright$ the induced graph $G_{An(\{v\})}$.
18:           $G' \leftarrow$ the mixed component of $G_{An(v)}$ containing v
19:           $A \leftarrow (A \cap (\text{vertices of } G')) \cup (\text{source nodes in } G')$
20:           **if** $\text{MaxFlow}(G'_{\text{flow}}(v, A)) = |pa(v)|$ **then**
21:              SolvedNode $\leftarrow$ SolvedNodes $\cup \{v\}$.
22:           **end if**
23:         **end if**
24:     **end for**
25: **until** SolvedNodes $= V$ or no change has occurred in the last iteration.
26: **Output:** "yes" if SolvedNodes $= V$, "no" otherwise.

---

in Algorithm 1. From this it follows that if the HTC-algorithm outputs "yes" then so will Algorithm 1.

It is straightforward to check that Algorithm 1 returns "yes" for the graph in Figure 1a and, thus, it remains only to argue that the time complexity is at most $\mathcal{O}(|V|^5)$. Note that the Max Flow algorithm for this problem has a running time of $\mathcal{O}(|V|^3)$ since G is acyclic, see Foygel et al. (2012a) for details. It is easy to see that this running time dominates on each iteration. Moreover, since at the end of each iteration through the $|V|$ nodes of the graph, the algorithm must either terminate or add at least one vertex to the set of solved nodes, it follows that there are at most $|V|^2$ iterations. We conclude that the maximum run time of the algorithm is $\mathcal{O}(|V|^2 \cdot |V|^3) = \mathcal{O}(|V|^5)$. □

Remark 3.8. One might expect that lines 18 to 22 in Algorithm 1 are superfluous. This is, however, false, and indeed we have found examples of graphs G on 10 nodes for which Algorithm 1 returns "yes" but the corresponding algorithm with lines 18 to 22 removed returns "no." As these examples are fairly large we have chosen to not display them here.

Algorithm 2 A procedure for generating random acyclic mixed graphs.

1: **Input:** A positive integer n and $0 < p, q < 1$
2: **Initialize:** A mixed graph $G = (V, D, B)$ with $V = \{1, ..., n\}$, $D = B = \varnothing$
3: Pick a random collection E of $n - 1$ bidirected edges so that (V, E) is a tree.
4: $B \leftarrow E$
5: **for** $1 \le i < j \le n$ **do**
6:     Add $i \leftrightarrow j$ to B with probability p
7: **end for**
8: **for** $1 \le i < j \le n$ **do**
9:     Add $i \to j$ to D with probability q
10: **end for**
11: **Output:** G

---

4. Computational Experiments

We now run a simulation study to examine the effect of applying Algorithm 1 to HTC-inconclusive graphs. All code is written in R, and we use the SEMID package to determine HTC-identifiability and HTC-unidentifiability (R Core Team, 2014; Foygel and Drton, 2013). For each combination of $n \in \{6, 8, 10, 12\}$, $p \in \{.1, .2, .3\}$, and $q \in \{.2, .3, .4, .5, .6\}$ we perform the following steps:

(i) Use Algorithm 2 with probability parameters p and q to generate random acyclic mixed graphs G with connected bidirected part on n nodes, until we have found 1000 graphs which are HTC-inconclusive.
(ii) For each of the 1000 HTC-inconclusive graphs G, use Algorithm 1 to test the generic identifiability of G.
(iii) Record the proportion of the 1000 graphs that are shown to be generically identifiable by Algorithm 1. Call this proportion $a_{n,p,q}$.

To summarize our findings we compute, for each pair (n, p), the average $b_{n,q} = \frac{1}{3} \sum_p a_{n,p,q}$. We then plot the values of $b_{n,q}$ in Figure 2. According to this figure, Algorithm 1 provides a modest increase in the number of graphs that are generically identifiable. This improvement is seen to be largest when q is large, that is, the directed part of the mixed graph is dense.

5. Conclusion

We have shown how the generic identifiability of a subgraph of a mixed graph G induced by an ancestral subset of vertices implies the generic identifiability of the corresponding edge coefficients in G (Theorem 3.3 and Corollary 3.4). We then provided, in Algorithm 1, one specific way of how to leverage this result by using the HTC of Foygel et al. (2012a) and the decomposition techniques of Tian (2005). Our new algorithm provides a modest strengthening of the HTC while not increasing the algorithmic complexity of the HTC-algorithm of Foygel et al. (2012a).

When saying above that Algorithm 1 constitutes one way of leveraging, we mean that the algorithm considers only certain ancestral subsets. While we do not have any examples to report, it is possible that there are acyclic mixed graphs for which Algorithm 1 does not return "yes" but which could be proven generically identifiable by a combination of Corollary 3.4, the Tian decomposition, and the HTC-algorithm. This said, it is not clear to us that all ancestral subsets can be considered in an
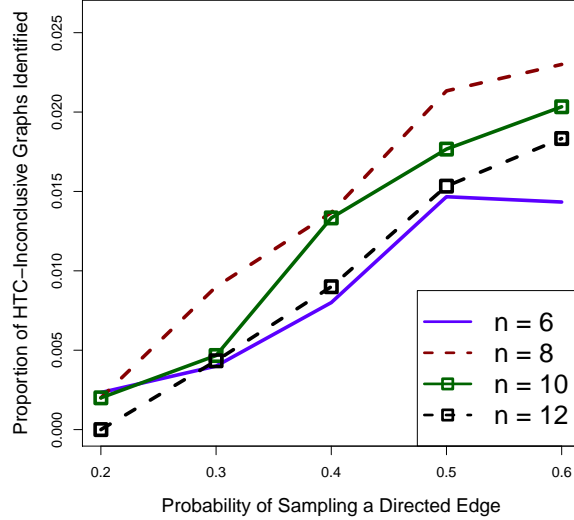
FIGURE 2. The average proportion of HTC-inconclusive graphs found to be generically identifiable by Algorithm 1.

algorithm with polynomial run time. Clarifying this issue would be an interesting topic for future work.

ACKNOWLEDGMENTS

REFERENCES

K. A. Bollen. *Structural equations with latent variables*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York, 1989. A Wiley-Interscience Publication.

C. Brito and J. Pearl. Graphical condition for identification in recursive SEM. In *Proceedings of the Twenty-Second Annual Conference on Uncertainty in Artificial Intelligence (UAI-06)*, pages 47–54. AUAI Press, 2006.

B. Chen, J. Tian, and J. Pearl. Testable implications of linear structural equations models. In C. E. Brodley and P. Stone, editors, *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, pages 2424–2430. AAAI Press, 2014.

M. Drton, R. Foygel, and S. Sullivant. Global identifiability of linear structural equation models. *Ann. Statist.*, 39(2):865–886, 2011.

R. Foygel and M. Drton. *SEMID: Identifiability of linear structural equation models*, 2013. R package version 0.1.

R. Foygel, J. Draisma, and M. Drton. Half-trek criterion for generic identifiability of linear structural equation models. *Ann. Statist.*, 40(3):1682–1713, 2012a.

R. Foygel, J. Draisma, and M. Drton. Supplement to half-trek criterion for generic identifiability of linear structural equation models. *Ann. Statist.*, 40(3), 2012b.

J. Pearl. *Causality.* Cambridge University Press, Cambridge, second edition, 2009. Models, reasoning, and inference.

R Core Team. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria, 2014.

I. Shpitser, R. J. Evans, T. S. Richardson, and J. M. Robins. Introduction to nested Markov models. *Behaviormetrika*, 41(1):3–39, 2014.

P. Spirtes, C. Glymour, and R. Scheines. *Causation, prediction, and search.* MIT press, Cambridge, MA, 2nd edition, 2000.

J. Tian. Identifying direct causal effects in linear models. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, pages 346–352. AAAI Press/The MIT Press, 2005.

S. Wright. Correlation and causation. *J. Agricultural Research*, 20:557–585, 1921.

S. Wright. The method of path coefficients. *Ann. Math. Statist.*, 5:161–215, 1934.

DEPARTMENT OF STATISTICS, UNIVERSITY OF WASHINGTON, SEATTLE, WA, U.S.A.
E-mail address: `md5@uw.edu`

DEPARTMENT OF STATISTICS, UNIVERSITY OF WASHINGTON, SEATTLE, WA, U.S.A.
E-mail address: `lucaw@uw.edu`